Theoretical Modelling of the Assembly of Small Icosahedral Viruses

Author: Marc Sadurní Parera
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: David Reguera López

Abstract: One of the key steps in the replication of a virus is the self-assembly of the capsid, which is a protein shell protecting the genetic material inside. The capsid plays an important role in the virus, and understanding its formation process would open the door to the development of new antiviral therapies. The aim of this work is to study theoretically the self-assembly of the smallest empty capsids, ignoring the presence of the genetic material. We formulate and compare a discrete description of viral assembly with a continuous approximation based on Classical Nucleation Theory. We have estimated the line tension involved in the formation of a partial icosahedral shell. Besides, we have solved numerically the Master equation using the estimated line tension in order to predict the nucleation rate for different concentrations. The results of our study show that Classical Nucleation Theory is valid even for very small viruses and can be a very powerful framework to analyze *in vitro* viral assembly experiments.

I. INTRODUCTION

Viruses are microscopic infectious agents which replicate inside the living cells of other organisms. In the simplest case they are composed by just an RNA or DNA chain packaged inside a protein shell, called *capsid*, which protects the genetic material. More than half of all known viruses have a capsid with icosahedral geometry. The triangulation number, T, is used to classify icosahedral viruses and gives information about the number of identical proteins in the capsid, which is $60T$ [1]. These proteins tend to cluster into aggregates of two, three, five or six of them, which are stable in solution and constitute the basic building blocks for capsid assembly. We will focus our study on the smallest icosahedral virus, corresponding to T=1, made of 60 identical proteins. More specifically, we will analyse the assembly of T=1 viruses like Minute Virus of Mice (MVM) [2], where the capsid building blocks (CBB) are trimers.

The self-assembly of the capsid is driven by the competition between different interactions. These include: the hydrophobic forces, due to the apolar patches of CBB; the electrostatic forces, due to the charged residues and the genetic material; the hydrogen bonds or the salt bridges. The outcome of the assembly depends very sensitively also on the temperature, the pH, and the salt concentration [3].

Different models have been proposed to try to explain virus capsid formation [1, 4–6]. In this work, we will present and compare a discrete description of viral assembly using a Master equation (ME) [1, 5–8], with a continuous approximation based on Classical Nucleation Theory (CNT) [3, 4, 7, 8]. We will particularize our analysis to a T=1 virus made of trimeric CBB. We will first estimate the nucleation barrier and the line tension for assembly. We will then implement a simulation to solve numerically the ME in order to predict the nucleation rate for different concentrations of CBB. The results of assembly at steady-state conditions have been compared to CNT, showing that the continuum approximation is valid even for small icosahedral viruses. Furthermore, these results could be very useful to analyze and predict viral assembly in *in vitro* experiments [2].

II. THEORETICAL MODELLING

A. Kinetic description of Viral Assembly

Virus capsid assembly occurs via a nucleation process. It is a phase transition between a metastable state, i.e., the CBB in solution, that helped by thermal fluctuations surmounts a free energy barrier, ending up in a stable state (the capsid).

The process of capsid assembly and the formation of intermediates that appear during capsid formation can be described using a Kinetic model. This model assumes that the assembly intermediates formed by the addition of n CBB can be considered as different "species" that go from 1 to q, where 1 corresponds to the individual subunits, and q is the number of CBB in a complete capsid. Moreover, it is assumed that the concentration of any species at time t just changes by the attachment or detachment of individual building blocks. Hence, the capsid formation process could be understood like a cascade of low-order reactions

$$[n](t) \underset{\alpha_{n+1}(t)}{\overset{\beta_n(t)}{\rightleftharpoons}} [n+1](t) \quad , \quad n = 2, ..., q-1 \qquad (1)$$

where $[n]$ is the concentration of partial capsids formed by n subunits at time t, and $\beta_n(t)$ and $\alpha_{n+1}(t)$ are the attachment and detachment rates, respectively. Generally, these rates could depend on time and a cluster size, n. The time variation of the concentration of a given species $[n]$ can be modelled by the following Master equation

(ME)

$$\frac{d[n]}{dt} = (\beta_{n-1}[n-1] - \alpha_n[n]) - (\beta_n[n] - \alpha_{n+1}[n+1]) \tag{2}$$

where it is assumed that all variables depend on time. This equation means that any partial capsid formed by n subunits can disassemble or assemble into $n-1$ or $n+1$ intermediate by the detachment or attachment of one CBB, respectively [1, 5, 6, 8].

To simplify the equations it is convenient to rewrite the master equation in terms of the flux as [7, 8]

$$\frac{d[n](t)}{dt} = J_{n-1}(t) - J_n(t) \tag{3}$$

where the flux is defined as

$$J_n(t) \equiv \beta_n(t)[n](t) - \alpha_{n+1}(t)[n+1](t). \tag{4}$$

Solving the set of master equations, one obtains the concentrations of all species. Nevertheless, in order to do that, first it is necessary to know the rates of attachment, $\beta_n(t)$, and detachment, $\alpha_n(t)$.

The rate of attachment $\beta_n(t)$ can be estimated using Smoluchowski theory of aggregation [8]

$$\beta_n(t) = bD_1 r_n[1](t) \tag{5}$$

where b is a geometric correction, D_1 is the diffusion coefficient of free subunits and r_n is the radius of the rim of the partial capsid.

However, the value of the detachment rate $\alpha_n(t)$ is difficult to estimate. This problem can be solved using the constrained equilibrium approximation, corresponding to $J_n = 0$. At equilibrium, the concentration of the different intermediates will be given by the standard Boltzmann distribution $[n]_{eq} = [1]_{eq} e^{-(\Delta G(n) - \Delta G(1))/k_B T}$, where $\Delta G(n)$ is the Gibbs free energy of formation of a partial capsid of n subunits. Using this approximation in Eq. 4 with $J_n = 0$, the detachment rate becomes [7–9]

$$\alpha_{n+1} = \beta_n e^{\left(\frac{\Delta G(n+1) - \Delta G(n)}{k_B T}\right)} \tag{6}$$

By inserting the previous equation into Eq. (4), it can be proved that the formula which describes the nucleation rate is [7, 8]

$$\sum_{n=1}^{q-1} \frac{J_n(t)}{\beta_n(t)[n]_{eq}} = \frac{[1](t)}{[1]_{eq}} - \frac{[q](t)}{[q]_{eq}} \tag{7}$$

This equation can be solved for different boundary conditions related to the concentrations of single CBB and complete capsids.

Let's first focus on the case where the total concentration of subunits, $N = \sum_{n=1}^{q} n[n](t)$, is constant during the assembly process. In this case, because the concentration of free subunits will vary, the nucleation rate, which is the number of capsids formed per unit of volume and time, will be extremely time-dependent, and the final state will correspond to equilibrium conditions. This situation can be implemented with the following boundary conditions

$$\frac{d[1]}{dt} = (-2\beta_1[1] + 2\alpha_2[2]) + ... + (-\beta_{q-1}[q-1] + \alpha_q[q])$$
$$\frac{d[q]}{dt} = \beta_{q-1}[q-1] - \alpha_q[q] \tag{8}$$

These equations reflect the fact that free subunits concentration can vary by the loss or gain of a single CBB in any partial capsid; and that the fully formed capsids cannot grow beyond size q, i.e., $\beta_q = 0$.

The second situation that we will analyze corresponds to steady-state conditions, characterized by $d[n]/dt = 0$. This means that all fluxes are the same, namely, $J_n = J_{n+1} \equiv J$, where J is the steady-state nucleation rate. The proper boundary conditions in this case are $[1](t)/[1]_{eq} = 1$, i.e., the steady-state concentration of free subunits is practically equal to the equilibrium one, and $[q](t)/[q]_{eq} = 0$. Thus, we are assuming that the steady-state concentration of fully formed capsids is vanishing small. These conditions are supposed to be a good approximation only for the initial stages of the process. Applying these conditions to Eq. (7), the steady-sate nucleation rate becomes [7–9]

$$J = \left(\sum_{n=1}^{q-1} \frac{1}{\beta_n(t)[n]_{eq}} \right)^{-1} \tag{9}$$

Finally, using Boltzmann's distribution for the equilibrium concentration, approximating the sum to an integral and applying the saddle-point approximation, the steady-state nucleation rate is obtained [7–9]

$$J = \beta(n^*)[1]_{eq} Z e^{-\frac{(\Delta G(n^*) - \Delta G(1))}{k_B T}} = \beta(n^*)[n^*]_{eq} Z \tag{10}$$

where $\beta(n^*)$ is the rate of attachment of individual CBB to the critical nucleus and $Z = \sqrt{\frac{1}{2\pi k_B T} |\partial^2 \Delta G / \partial n^2|_{n=n^*}}$ is the Zeldovich factor associated with the local curvature at the maximum of the barrier. Eq. 10 is a good approximation when the nucleation barrier is big enough.

All these equations depend on the Gibbs free energy of capsid formation, which can be evaluated using CNT, as described below.

B. Classical Nucleation Theory (CNT)

The CNT develops a physical model for the Gibbs free energy of partial capsids formation. This energy corresponds to the energy difference between the CBB in an incomplete shell and in solution. It contains two main terms. A "bulk" term, reflecting the energy gain due to

the favourable binding energy; and a line tension correction that appears for incomplete capsids. This correction is necessary since the CBB at the rim of a partial shell have less contacts than in a complete shell. Therefore, the formation of intermediate states entails an energetic cost reflected in an energy barrier. The Gibbs free energy of formation of a capsid with n building blocks is given by [3, 4, 7, 8]

$$\Delta G(n) = n\Delta\mu + \sigma l = n(\mu_c - \mu_s) + \sigma l \tag{11}$$

where $\Delta\mu = -K_B T \ln\left(\frac{[1]}{c^*}\right)$ is the chemical potential difference between CBB in the capsid and in solution, $[1]$ is the concentration of free monomers, c^* is the threshold concentration, l is the contour length of the rim and σ is the line tension reflecting the energy cost per unit of length associated with the rim.

If the concentration of free subunits exceeds the threshold concentration c^* the formation of a shell is favourable since the chemical potential of proteins in a fully formed capsid, μ_c, becomes smaller than in solution, μ_s. The difference is the driving force for capsid formation. If $[1] < c^*$ the formation of a capsid is unfavourable.

The free energy of a partial shell can be evaluated with a quasi-continuum approach assuming that the capsid can be considered as a spherical cap of radius R made of n subunits (see FIG. 1 (a)). From simple geometry, it can be proved that $l = 2\pi R \sin\theta$ and $n = q(1-\cos\theta)/2$. Hence, l can be expressed as a function of n as [7, 8]

$$l = \frac{4\pi R}{q}\sqrt{n(q-n)}. \tag{12}$$

Using this approximation, the Gibbs free energy in terms of the number of subunits n in a partially formed capsid is

$$\Delta G(n) = n\Delta\mu + a\sqrt{n(q-n)} \tag{13}$$

where $a = \frac{4\pi R\sigma}{q}$ is a measure of the line tension energy.

The competition between the surface and the line tension terms give rise to the nucleation barrier

$$\Delta G(n^*) = \frac{q}{2}\left(\Delta\mu + \sqrt{\Delta\mu^2 + a^2}\right) \tag{14}$$

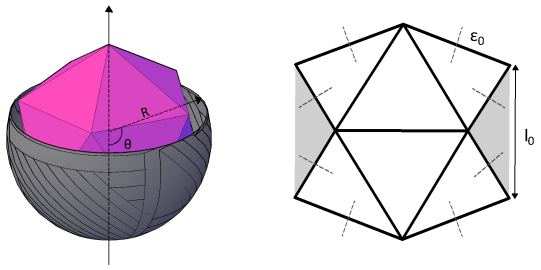
where the location of the maximum

$$n^* = \frac{q}{2}\left(1 + \frac{\Delta\mu}{\sqrt{\Delta\mu^2 + a^2}}\right) \tag{15}$$

corresponds to the critical cluster [7, 8].

For $n < n^*$, the Gibbs free energy increases with the number of CBB. Therefore, small partial capsids tend to disassemble back into free subunits. However, for partial capsids with $n > n^*$, the free energy decreases by the addition of CBB. As a consequence, they will tend to grow up spontaneously towards a complete capsid. Then, n^* is the critical size that the partial shells have to reach to activate the formation of full capsids. For that reason, n^* called the *nucleus* or *embryo* of the process.

In this study, we will particularize the ME and CNT to analyze the assembly of a T=1 icosahedron made of 20 trimeric CBB.



(a) (b)

FIG. 1: (a) Continuous approximation of a partially formed icosahedron by a spherical cap. R is the radius of the capsid and θ is the angle which characterizes the degree of completion. (b) Estimate of the line tension of an equilateral triangular lattice in 3-D by cutting perpendicularly to the 2-fold symmetry. The lost triangles are represented in gray. By removing these triangles, 8 contacts of energy ϵ_0 (dashed lines) are broken per 6 units of length l_0.

III. RESULTS AND DISCUSSION

A. Characterization of the line tension

One of the main ingredients of CNT is the line tension correlated to the formation of a partial capsid. In order to determine it, the boundary energy, $E_b(n, R)$, of a partial shell has been computed for the discrete and continuous cases.

As a first approximation, the $E_b(n, R)$ in the discrete theory is given by the total energy in a partial shell less the energy of a completed shell as [10]

$$E_b(n, R) = E(n, R) - \mu_e n \tag{16}$$

where $E(n, R) = -C\epsilon_0$ is the total energy of the shell, estimated as the number of contacts C between CBB times the binding energy per contact $-\epsilon_0$, and μ_e is the chemical potential at equilibrium, corresponding to the energy per CBB in a complete capsid. We have assumed that the elastic and binding energies of the shell are negligible. In an icosahedral virus composed by 20 equilateral triangular equilateral faces, each contact is shared by two CBB, and each CBB has three neighbours. Thus, $\mu_e = -\frac{3}{2}\epsilon_0$ (see FIG. 1 (b)).

n	Model	E	E_b	n	Model	E	E_b
1		0	1.5	11		13	3.5
2		1	2	12		15	3
3		2	2.5	13		16	3.5
4		3	3	14		18	3
5		5	2.5	15		20	2.5
6		6	3	16		21	3
7		7	3.5	17		23	2.5
8		9	3	18		25	2
9		10	3.5	19		27	1.5
10		12	3	20		30	0

TABLE I: Assembly intermediates corresponding to the most stable path. All energies are in units of ϵ_0. The degeneracy of each step is not considered.

The energy $E(n, R)$ has been estimated by counting the number of contacts in a shell growing following the the optimal path. This means for a given n, the fragment of an icosahedron having the highest number of contacts, C. TABLE I includes all assembly intermediates and parameters describing the most stable path.

In the context of a CNT, it was shown that the boundary energy is given by

$$E_b(n, R) = \sigma l = \frac{4\pi R\sigma}{q}\sqrt{n(q-n)}. \tag{17}$$

The effective radius $R = \sqrt{\frac{5\sqrt{3}}{4\pi}}l_0$ can be found as the radius of a sphere having the same area of an icosahedron, i.e., $Aq = 4\pi R^2$, where $A = \frac{\sqrt{3}}{4}l_0$ is the area of each CBB, considered as an equilateral triangle of side length l_0.

FIG 2 represents the boundary energy as a function of n. The figure also includes the fit to the functional dependence predicted by CNT, $f(n) = a\sqrt{n(20-n)}$. The line tension obtained from the fit is

$$\sigma = 0.64\frac{\epsilon_0}{l_0}. \tag{18}$$

Alternatively, the line tension can be estimated as one half of the energy of the lost contacts per unit of length when cutting the icosahedron along the directions corresponding to 2-fold, 3-fold and 5-fold symmetry axes. The result which adjusts better with Eq. 18 is obtained by

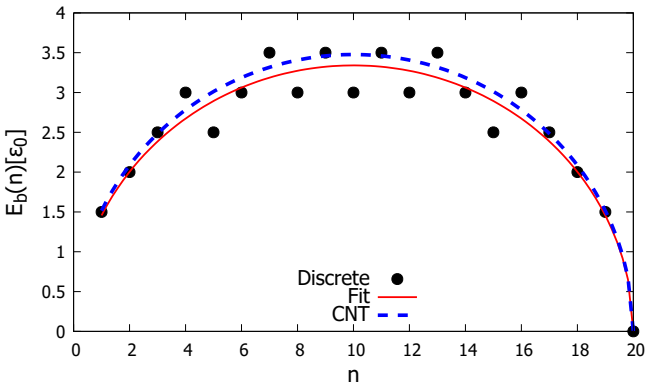


FIG. 2: Boundary energy, $E_b(n, R)$, of a partial rigid capsid, in units of ϵ_0, as a function of the number of subunits n, for 3 cases: discrete (black points), fit (red line) and CNT (blue dashed line).

cutting perpendicularly to the 2-fold direction (see FIG. 1 (b)). In this case, 8 contacts are lost per $6l_0$. Hence, the estimated line tension is

$$\sigma_{2-fold} = \frac{8}{12}\frac{\epsilon_0}{l_0} = \frac{2}{3}\frac{\epsilon_0}{l_0}. \tag{19}$$

This last result is also shown as the blue dashed line in FIG 2. The figure shows that CNT provides a remarkably good quantitative approximation to the barrier for the self-assembly of even a small T=1 virus.

B. Numerical Simulations

In this section, the ME (Eq. 2) has been solved using a simple Euler method implemented in a Fortran code for two different boundary conditions.

In order to make the results valid for any type of virus, the simulations have been performed using reduced units in terms of a characteristic unit of length l_0, which represents the triangular equilateral side; a unit of energy ϵ_0, which is the binding energy between two CBB, and diffusion coefficient D_0. Using these reduced units, the unit of time becomes $t_0 = l_0^2/D_0$. The typical value of the binding energy between capsid proteins is $\epsilon_0 = 5k_BT$ [11]. Thus, in our calculations we have used a reduced temperature of $k_BT = 0.2\epsilon_0$. All simulations are solved taking $c^* = 1$ in reduced units, using the line tension estimated before, and assuming that the initial state is composed only by free subunits.

FIG. 3 presents the results of the numerical simulation for fixed N and time-dependent conditions (Eq. 8). The figure shows the free subunits, the half formed capsids, and the fully formed shells concentrations. In this situation the system converge to the equilibrium state at long times, where most of the concentrations are free subunits and completed capsids. Besides, the lag time,

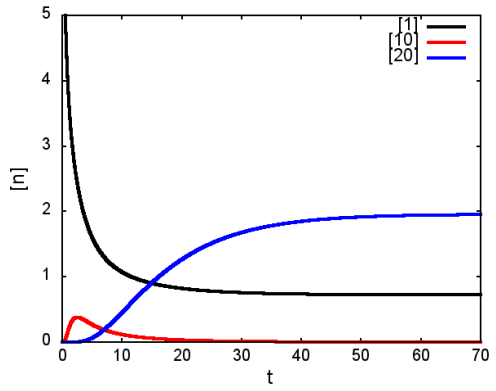


FIG. 3: Individual CBB, half formed capsids, and fully formed shells concentrations as a function of time. The simulation is solved with initial concentration $[1](t=0)=40$.

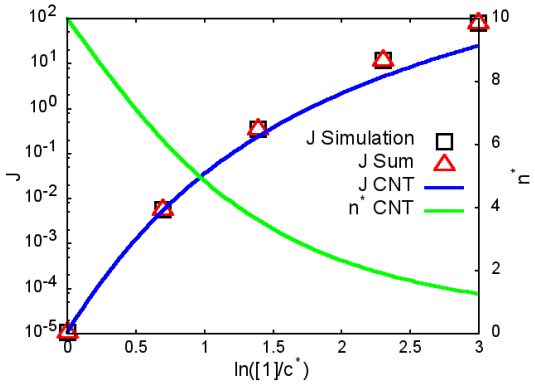


FIG. 4: Steady nucleation rates and n^* as a function of $ln([1]/c^*)$. The nucleation rates are computed numerically (black squares), using Eq. 9 (red triangles) and Eq. 10 (blue line). The concentrations of subunits, in reduced units, from left to right are: 1, 2, 4, 10 and 20, respectively. The critical size predicted by CNT is shown as a green line.

i.e., the time to formation the first completed capsid, can be seen.

FIG. 4 is the result of simulating capsid assembly at steady-state conditions. This figure unquestionably shows that CNT approximates the numerical steady nucleation rates when the energy barrier is high, i.e., when the concentration of subunits is similar to c^*. Instead, for a smaller barrier, i.e., when the concentration of CBB is much larger than c^*, CNT gives a modest deviation from the discrete rates. Moreover, the critical nucleus n^* prediction is represented in order to compare with the viral assembly in the experiments [2].

IV. CONCLUSIONS

In this work we have presented a theoretical description of the assembly of a T=1 virus. The line tension of this virus was estimated considering the optimal path of the self-assembly. Furthermore, the ME was solved for different conditions obtaining viral nucleation concentrations and rates. CNT approximates very well the discrete results when the energy barrier is high, whereas for low energy barrier the CNT deviates slightly.

The model presented here is applicable for any T=1 virus. Thus, it would be interesting to apply it to a specific virus, such as the Minute Virus of Mice (MVM) [2], in order to predict the self-assembly of these viruses experimentally and estimate several parameters such as the critical concentration and the binding energy.

Acknowledgments

I would like to thank my advisor David Reguera for the whole dedication, enthusiasm and reliance disposed on me, which has successfully boost me to develop this scientific article. Also my most sincere thanks to the whole support received from my family and friends during my studies.

[1] A. Zlotnick. *To Build a Virus Capsid.* J. Mol. Biol. **241**, 59-67 (1994)
[2] M. Medrano, M.A. Fuertes, A. Valbuena, P.J.P. Carrillo, A. Rodríguez-Huete, M.G. Mateu. *Imaging and Quantitation of a Succession of Transient Intermediates Reveal the Reversible Self-Assembly Pathway of a Simple Icosahedral Virus Capsid.* J. Am. Chem. Soc. **138**, 15385-15396 (2016)
[3] A. Luque, D. Reguera. *Structure and Physics of Viruses: Chapter 19.* Editor: M.G. Mateu. SCBI **68** (2013)
[4] R. Zandi, P. Schoot, D. Reguera, W. Kegel, H. Reiss. *Classical Nucleation Theory of Virus Capsids.* Biophys. J. **90**, 1939-1948 (2006)
[5] T. Keef, C. Micheletti, R. Twarock. *Master equation approach to the assembly of viral capsids.* J. Theor. Biol. **242**, 713-721 (2006)
[6] D. Endres, A. Zlotnick. *Model-Based Analysis of Assembly Kinetics for Virus Capsids or Other Spherical Polymers.* Biophys. J. **83**, 1217-1230 (2002)
[7] J. Wedekind, A. Luque, D. Reguera. *Nucleation theory.* Pre-print (2009)
[8] A. Luque. *Structure, mechanical properties, and self-assembly of viral capsids.* PhD thesis, UB (2011)
[9] D.T. Wu. *Nucleation theory.* Solid State Physics **90**, 39-93 (1997)
[10] A. Luque, D. Reguera, A. Morozov, J. Rudnick, R. Bruinsma. *Physics of shell assembly: Line tension, hole implosion, and closure catastrophe.* J. Chem. Phys. **136**, 184507 (2012)
[11] J. Wedekind, J. Wölk, D. Reguera, R. Strey. *Nucleation rate isotherms of argon from molecular dynamics simulations.* J. Chem. Phys. **127**, 154515 (2007)